In this project, students will process raw data on reported Covid-19confirmed cases around the globe. The project’s objective is to train students on a real-life scenario: how to use Python to retrieve vital information from raw MS Excel or CSV data.

Instructions:

1. Platform: You may use any Python IDE for this project. However, I suggest you use Jupyter Notebooks. For that, install an aconcaon your computer(https://docs.anaconda.com/anaconda/install/);

it comes with most packages you will need, including Jupyter. Follow the instructions on the web page. You may install any package or import any Python module you deem necessary for the task.

1. Data: A set of three .csv data files are provided:

a. population.csv contains human population in all countries and territories in the world;

b. countries.csv contains data on countries classification per geographical region and continent;

c. time\_series\_covid19\_confirmed\_global.csv contains officially reported Covid-19 infection cases per country,per province, state, or territories.

1. Tasks

1.(20 points) Create a Python module called xy\_covid19\_module.py(x and y are your name initials), which will contain all your functions. The main function to be called to execute all the tasks in this project should be named xy\_covid19\_main\_func(),which receives the three files in II as arguments and calls your other functions to perform all the tasks in this project.One should be able to import your module and call the xy\_covid19\_main\_func()function by passing the three files to it as arguments.

Task

2.(20 points) Combine useful data from different files: white one or more Python functions to automatically create a file named xy\_per\_country\_data.csv(x and y are your name initials)file with the following columns:

a. “country” according to the list of countries in column 1 of population.csv;

b. “population” (population of that country)

c. “mm/dd/yyy” between 3/01/2020 and 3/31/2020(only for the month of March 2020)with the daily reported cases (just like in the time\_series\_covid-19\_confirmed\_global.csv file). Note that for countries

that reported numbers by state, province,or territory, you need to calculate the sum of those numbers.

d.“geographical\_region”

e.“continent”

f.“#\_cases\_per\_1000”the number of cases per thousand in habitants

g. “population\_density”

h.“cases\_per\_km2”

i.“latitude”

Task

3.(40 points)Answer questions: Write a function named xy\_answers\_func() to Createa file named xy\_answers.txt to answer the following questions, one answer per line.

a. Number of countries?

b. Country w/ most casesin3/ 2020 ? :

c. Country w/least cases in 3/2020?:

d. Country w/ no cases in 3/2020?: (if multiple countries, separate them by a comma)

e. Most densely populated country?:

f. Least densely populated country?:

g. For every continent, provide:

i. Country w/ most cases in 3/2020?:

ii. Country w/least cases in 3/2020?:

iii. Country w/ no cases in 3/2020?: (if multiple countries, separate them by a comma)

h. If it is assumed the hypothesis “the higher the population density, the higher the risk for the country.” Name:

i. The top 10at risk countries in the world?:

ii. The top 10least risky countries?:

iii. Is the hypothesis true (yes/no)?:

iv. Why that answer?:

i. If it is assumed the hypothesis “country that are crossed by the equator are not affected by Covid-19.” Name:

i. Countries of latitude between -2 and 2 degrees?:

ii. For each country in “i,”what is:1.the rank by number of cases in 3/2020?:2.the rank by“#\_cases\_per\_1000”?:

iii. Is the hypothesis true?

Task

4.(20 points) Plotting: Create functions to plot the followings:

a. A line chart to show the daily reported cases for the most affected (high est total number of cases in March 2020) country in each geographical region, with a legend displaying the country and its region;

b. A bar chart to show the total number of cases in each geographical region with a legend;

c. A pie chart to show the percentage of cases in each geographical region with regard to the total number of global cases.

5.Submission:

Submit the ONLY the xy\_covid19\_module.py file.